☐ | ⎡ Generate Collection ⎤ | ⎡ Print ⎤

L4: Entry 2 of 3   File: USPT   Sep 20, 2005

DOCUMENT-IDENTIFIER: US 6947947 B2
TITLE: Method for adding metadata to data

Brief Summary Text (15):
In an exemplary embodiment of the invention, a data stream is captured, data in the captured
stream are identified, and then the identified data are mapped to a file structure, a schema,
or a taxonomy. In exemplary embodiments of the invention, the output data stream is a data
stream to a display screen, a memory, a hard drive, a CD ROM drive, a floppy disk drive, or a
printer. The output data stream can be conveyed through serial or parallel ports (including
Universal Serial Bus or "USB", FireWire.TM.), via wireless interfaces, and so forth. In other
exemplary embodiments of the invention, the identified data are mapped to an XBRL (eXtensible
Business Reporting Language) taxonomy, a spreadsheet, a database, or a flat file.

Brief Summary Text (20):
Exemplary embodiments of the invention include a synonym dictionary that includes synonyms of
known labels or terms, or synonymous links between labels and/or terms, to facilitate automatic
or user-assisted mapping. The dictionary can include terms that are not part of a taxonomy or
schema such as an XML taxonomy, but that are synonymously related to terms in a taxonomy,
schema, etc. In an exemplary embodiment of the invention, the synonym dictionary includes
foreign languages, so that a label or datum can be mapped from one language into another
language. In an exemplary embodiment of the invention, currency values are identified in the
data stream, and are converted to corresponding values in different currencies (e.g., from yen
to dollars) based on a known or designated exchange rate. In accordance with an embodiment of
the invention, the mapping process converts data from one standard to another, for example from
U.S. GAAP (Generally Accepted Accounting Principles) to International GAAP. In accordance with
an embodiment of the invention, the mapping process includes replacing labels corresponding to
identified data, with other labels, for example where minimizing file size is important.

Brief Summary Text (21):
In accordance with an embodiment of the invention, data output from a first computer platform
or system can be automatically converted by a software module on the first platform, from a
first format into an intermediate format, transferred to a second platform or system, and then
converted from the intermediate format into a second format by a second software module on the
second platform. For example, the intermediate format can be an XML taxonomy, and the software
modules can effectively "translate" so that data can be transparently exchanged between the two
platforms regardless of whether the first and second formats are compatible or known to each of
the two platforms. The intermediate format can also be encrypted, e.g. for secure transfer.

Detailed Description Text (2):
In accordance with an embodiment of the invention shown in FIG. 1A, a data stream is captured
in step 150, data in the captured data stream are identified in step 152, and then in step 154
the identified data are mapped to a file structure, a schema, or a taxonomy. The output data
stream is a data stream to a display screen, a memory, a hard drive, a CD ROM drive, a floppy
disk drive, or a printer. The output data stream can be conveyed within a computer, through
serial or parallel ports (including Universal Serial Bus or "USB", FireWire.TM., etc.), via
wireless interfaces, and so forth, and can be captured via duplication or redirection, at any
point along the conveyance, via software and/or hardware mechanisms. The identified data are
mapped to an XBRL (eXtensible Business Reporting Language) taxonomy, a spreadsheet, a database,
an XML (eXtensible Markup Language) taxonomy, a standard (e.g., U.S. GAAP or International
GAAP), or a flat file. When the identified data are mapped to a flat file, a specification or
"data definition" file can also be generated to indicate the meaning or character of
information at different locations in the flat file (e.g., in different columns, at different
locations within a given text string, etc.), and to optionally indicate delimiters (e.g. tabs,

commas, spaces, semicolons, etc.) between discrete elements of information or groups of information in the flat file. The flat file and an accompanying data definition can, for example, be generated in accordance with known techniques and formats relating to flat files.

Detailed Description Text (3):
The embodiment shown in FIG. 1A, can be implemented as shown in FIG. 1B. In accordance with an exemplary embodiment of the invention illustrated in FIG. 1B, a method for adding labels to data includes a) identifying data in an electronically represented file, b) selecting labels that correspond to metadata in the identified data, based on a list associating labels with metadata, and c) adding the selected labels into the electronically represented file to label the metadata and/or elements in the identified data associated with the metadata. The labels include information about the data and are defined in one or more taxonomies. In the context of the present application, "metadata" or "meta information" is data about data, or information that describes other information. In this example the metadata in the identified data identifies or describes other data elements within the identified data, and can include for example text strings, various control characters (e.g., various ASCII control characters), and so forth For example, metadata in the captured data stream or file can be used to identify the data to which the metadata refer, and then additional metadata referring to the identified data can be added to the captured data stream or file. For example, the list can contain labels from multiple taxonomies, standards, and so forth, including words from languages, link synonymous or related labels. When a label from a first taxonomy, etc. is recognized in the captured data stream or file, the data element it labels can also be further labeled with a corresponding label from a second, different taxonomy, standard, etc. Thus a computer program that recognizes the second taxonomy but not the first, will now be able to use or recognize and organize the information in the data stream or file. A new, transformed data stream or file can be formed by adding the new labels for the second taxonomy, and optionally removing the old labels from the first taxonomy (or standard, schema, etc.).

Detailed Description Text (4):
In the event the list does not associate a label with metadata in the identified data, a user can be prompted to select a label corresponding to the metadata. The association indicated by the user's selection, can then be added to the list associating labels with metadata. Preferably the labels are consistent with XML (eXtensible Markup Language), and also conform to an XBRL (eXtensible Business Reporting Language) specification. Of course, the labels can also be consistent with data formats for spreadsheets, relational databases, and other file structures or schemas or standards.

Detailed Description Text (6):
The transformation program can be independent and separate from the target program. The transformation program can also be entirely resident on the same computer or system as the target program, or can be remotely located on a different system, or distributed among different systems. The transformation module can be a single module, or a plurality of cooperating modules. A list and/or synonym dictionary that the transformation program or module (s) use to identify metadata and add corresponding metadata, can be stored as a data file separately from the program or module(s), and can be stored or accessed remotely, for example via an Internet web server.

Detailed Description Text (8):
The transformation programs on the provider and receiver machines can be identical and both capable of receiving, transferring and mapping data, or can have different capabilities. For example, the transformation programs can be configured to handle an intermediate format so that the transformation program at the information provider would map the data to an intermediate format, and transfer the data in the intermediate format to the transformation program on the receiver machine. The receiver machine would map the data from the intermediate format to another format useful on the receiver machine (or as desired by a user). The programs could be different versions, so that the transformation program recognizes more formats than the transformation program at the receiver machine and thus can map more formats to or from the intermediate format. In addition or as an alternative, the transformation program on the receiver machine can be configured or featured to only map the data out of the intermediate format to another format, without being able to map data into the intermediate format in much the same way that Adobe Acrobat.TM. Readers can open and view, but not create, .pdf files. The transformation programs can also be configured to operate automatically without user intervention. For example, the transformation program on the provider machine can automatically

transfer data in response to a request from the transformation program on the receiver machine, subject for example to rules or requirements (e.g., a user's prior approval to allow public access to information on the provider machine) in place on the provider machine. The provider and receiver machines can communicate via the Internet. For example, the provider machine can interface the Internet or function as a web server, and the receiver machine can interface the Internet or function as a web browser. Also, the intermediate format can be encrypted, and can be decrypted at the receiver machine in a fashion transparent to a user of the receiver machine. For example, the encryption/decryption mechanism can be a proprietary function of the transformation programs.

Detailed Description Text (20):
Exemplary embodiments of the invention include a synonym dictionary that includes synonyms of known labels or terms, or synonymous links between labels and/or terms, to facilitate automatic or user-assisted mapping. For example, where a known label in a standard, schema or taxonomy to which captured data stream or file is being mapped is "Sales", the dictionary can include synonyms such as "Fees" and "Revenues" so that when the synonyms are identified in the captured data stream the datum they refer to will be mapped appropriately to (or labeled with) the label "Sales". The synonym dictionary can be incorporated within the list associating data and metadata. The dictionary can include terms that are not part of a taxonomy or schema such as an XML taxonomy, but that are synonymously related to terms in a taxonomy, schema, etc. In an exemplary embodiment of the invention, the synonym dictionary includes foreign languages, so that a label or datum can be mapped from one language into another language.

CLAIMS:

60. A computer implemented method for adding metadata to a collection of data and first metadata wherein the first metadata are associated with the data, the method comprising: identifying data in the collection based on the first metadata and one or more locations of the data and/or the first metadata in the collection; and adding second metadata to the collection based on the identified data; wherein the second metadata map the identified data to at least one of a file structure, a schema, and a taxonomy.

132. A machine readable medium comprising a computer program for causing a computation device to perform: in a collection of data and first metadata wherein the first metadata are associated with the data, identifying data in the collection based on the first metadata and one or more locations of the data and/or the first metadata in the collection; and adding second metadata to the collection based on the identified data; wherein the second metadata map the identified data to at least one of a file structure, a schema, and a taxonomy.

183. A system comprising a computation device arranged to perform: in a collection of data and first metadata wherein the first metadata are associated with the data, identifying data in the collection based on the first metadata and one or more locations of the data and/or the first metadata in the collection; and adding second metadata to the collection based on the identified data; wherein the second metadata map the identified data to at least one of a file structure, a schema, and a taxonomy.

Freeform Search

Database:

```
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins
```

Term:

```
L3 and (target near system)
```

Display: `50` Documents in **Display Format:** `FRO` **Starting with Number** `1`

Generate: ○ **Hit List** ⊙ **Hit Count** ○ **Side by Side** ○ **Image**

| Search | Clear | Interrupt |

Search History

DATE: Friday, February 01, 2008   **Purge Queries**   Printable Copy   Create Case

Set Name	Query	Hit Count	Set Name
side by side			result set
	DB=USPT; PLUR=YES; OP=OR		
L4	L3 and (target near system)	3	L4
L3	L1 and ((intermediate or third) near format) and schema	31	L3
L2	L1 and ((intermediate or third) near format)	577	L2
L1	((first and second) near format)	3428	L1

END OF SEARCH HISTORY

Freeform Search

Database:
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Term:
L3 and (target near system)

Display: 50 Documents in **Display Format:** FRO **Starting with Number** 1

Generate: ○ **Hit List** ⦿ **Hit Count** ○ **Side by Side** ○ **Image**

Search   Clear   Interrupt

Search History

DATE: Friday, February 01, 2008   **Purge Queries**   ·Printable Copy   Create Case

Set Name	Query	Hit Count	Set Name
side by side			result set
	DB=USPT; PLUR=YES; OP=OR		
L4	L3 and (target near system)	3	L4
L3	L1 and ((intermediate or third) near format) and schema	31	L3
L2	L1 and ((intermediate or third) near format)	577	L2
L1	((first and second) near format)	3428	L1

END OF SEARCH HISTORY